# Adding levels of complexity enhances robustness and evolvability in a multilevel genotype–phenotype map

Catalán, Pablo ; Wagner, Andreas ; Manrubia, Susanna ; Cuesta, José A

Abstract: Robustness and evolvability are the main properties that account for the stability and accessibility of phenotypes. They have been studied in a number of computational genotype–phenotype maps. In this paper, we study a metabolic genotype–phenotype map defined in toyLIFE, a multilevel computational model that represents a simplified cellular biology. toyLIFE includes several levels of phenotypic expression, from proteins to regulatory networks to metabolism. Our results show that toyLIFE shares many similarities with other seemingly unrelated computational genotype–phenotype maps. Thus, toyLIFE shows a high degeneracy in the mapping from genotypes to phenotypes, as well as a highly skewed distribution of phenotypic abundances. The neutral networks associated with abundant phenotypes are highly navigable, and common phenotypes are close to each other in genotype space. All of these properties are remarkable, as toyLIFE is built on a version of the HP protein-folding model that is neither robust nor evolvable: phenotypes cannot be mutually accessed through point mutations. In addition, both robustness and evolvability increase with the number of genes in a genotype. Therefore, our results suggest that adding levels of complexity to the mapping of genotypes to phenotypes and increasing genome size enhances both these properties.

DOI: https://doi.org/10.1098/rsif.2017.0516

# Adding levels of complexity enhances robustness and evolvability in a multi-level genotype-phenotype map

Pablo Catalán, Andreas Wagner, Susanna Manrubia and José A. Cuesta

## Abstract

Robustness and evolvability are the main properties that account for the stability and accessibility of phenotypes. They have been studied in a number of computational genotype-phenotype maps. In this paper we study a metabolic genotype-phenotype map defined in toyLIFE, a multi-level computational model that represents a simplified cellular biology. toyLIFE includes several levels of phenotypic expression, from proteins to regulatory networks to metabolism. Our results show that toyLIFE shares many similarities with other seemingly unrelated computational genotype-phenotype maps. Thus, toyLIFE shows a high degeneracy in the mapping from genotypes to phenotypes, as well as a highly skewed distribution of phenotypic abundances. The neutral networks associated with abundant phenotypes are highly navigable, and common phenotypes are close to each other in genotype space. All of these properties are remarkable, as toyLIFE is built on a version of the HP protein folding model that is neither robust nor evolvable: phenotypes cannot be mutually accessed through point mutations. In addition, both robustness and evolvability increase with the number of genes in a genotype. Therefore, our results suggest that adding levels of complexity to the mapping of genotypes to phenotype enhances both these properties.

## 1 Introduction

Classical evolutionary models do not account for the robustness and evolvability of phenotypes [1]. They thus fail to explain some evolutionary phenomena, such as punctuated equilibria [2, 3], constrained evolution [4], or the origins of novelty [5, 6]. In recent years, several research groups have tried to understand this important question by studying computational mappings of molecular genotypes to phenotypes. Some of these maps try to remain faithful to biological phenomena, such as RNA secondary structure [7–14], protein secondary structure [15–20], gene regulatory networks [21–24] and metabolic networks [25–28]. More abstract models have also been developed, such as the polyomino [29–31] and toyLIFE [32], as well as the Fibonacci map [33], and simple combinatorial maps [34]. The robustness and evolvability of phenotypes has also been recently studied for the artificial life AVIDA system [35].

Even though all these models focus on different aspects of molecular biology, all of them share some common properties. First, the mapping from genotype to phenotype is highly degenerate: many genotypes encode the same phenotype. Additionally, phenotype abundance (the number of genotypes encoding it) is not evenly distributed: most phenotypes are rare, while a few of them are extremely abundant. The probability density function associated with phenotype abundance is a log-normal distribution for a wide variety of models [14, 34], although in some cases it has been found to be a power law [33, 34]. This implies that rare phenotypes will not play a central role in evolution: they are hard to find in a genotype space that is filled with abundant phenotypes [12, 14, 36]. This degeneracy is usually accompanied by the formation of neutral networks, networks of genotypes encoding the same phenotype, in which two genotypes are connected if they differ in one point mutation —i.e. if the strings representing each genotype differ in one letter [7, 16, 21, 25, 37]. The degree of a node in such a neutral network —the number of neutral neighbors it has— is called genotypic robustness. It is usually normalised by the total number of neighbors in genotype space, and thus represents the fraction of neighbors with the same phenotype [5]. Inside a particular neutral network, the degree distribution is highly heterogeneous, but usually unimodal [13, 38]. Additionally, in RNA, the average degree of a neutral network grows linearly with the logarithm of its abundance [13] —this positive correlation is also observed in the polyomino model [30], the HP model [31] and simple genotype-phenotype maps [34], and is suggested by empirical data [39]. Neutral networks of abundant phenotypes percolate genotype space: they contain genotypes that share almost no letters [5, 7, 21]. Conversely, most abundant phenotypes are easily accessed from each

other: traversing a neutral network, many phenotypes can be found at its boundary [5, 21, 25]. Moreover, abundant phenotypes are typically found just a few mutations away from a random genotype [5, 9, 30] —i.e. they are highly evolvable. This means that these phenotypes are easily accessible from any other phenotype, so that the search for new phenotypes among abundant ones is a fast evolutionary process.

In [32] we presented toyLIFE, a multi-level model of a genotype-phenotype map (see Supplementary Material, Section 1 for a summary of toyLIFE's definition). toyLIFE includes genes, proteins, regulatory networks and a simple metabolism. Genes are binary strings of fixed length, divided in two regions —a promoter and a coding region. The coding region is translated into a sequence of 16 amino-acids that folds on a $4 \times 4$ lattice, following the rules of the HP protein-folding model [15–17]. Once folded, we only distinguish proteins by their perimeter and folding energy — note that this definition of folded protein is different from other versions of the HP model. In toyLIFE, there are $2,710$ different proteins (Supplementary Figure 3), most of them obtained from more than one sequence. However, there are no neutral mutations: every change in a coding region will result (most of the time) in a non-folding protein or (more rarely) in a different functional protein. This is very different from what has been observed in other versions of the HP model [17–20]. Proteins in toyLIFE interact with each other to form dimers, and both proteins and dimers regulate gene expression and interact with metabolites. The phenotype in toyLIFE can be defined in multiple ways. Here we will focus on a metabolic definition of phenotype, similar to the one presented in [25, 40]: the set of metabolites that a genotype can metabolise (Supplementary Material, Section 2). toyLIFE is, to our knowledge, the only multi-level genotype-phenotype map, incorporating genetic dynamics, protein folding, regulatory networks and metabolism.

In this paper we investigate the characteristics of the metabolic genotype-phenotype map of toyLIFE, showing that toyLIFE shares many properties with other computational genotype-phenotype maps. toyLIFE is highly robust and evolvable, even though it is built from a non-robust model of protein folding. These properties arise and increase as new genes are added to the genome, suggesting an important role for complexity on robustness and evolvability.

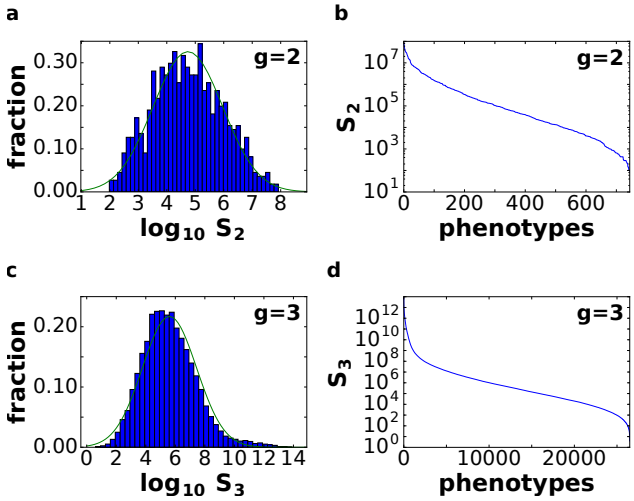


Figure 1: **Degeneracy and asymmetry in the genotype-phenotype map of toyLIFE.** **(a)** The distribution of the number of genotypes per phenotype —phenotype abundance— for $g = 2$ ($S_2$) follows a log-normal distribution, with probability density function $f(x) = (x\sigma\sqrt{2\pi})^{-1} \exp(-(\log x - \mu)^2/2\sigma^2)$, where $\mu$ is the mean and $\sigma$ is the standard deviation of the normally distributed logarithm of the variable. Here $\mu = 4.742$ and $\sigma = 1.224$ (empirically obtained from the log-transformed abundance distribution). **(b)** The rank distribution shows a long tail of small phenotypes. **(c)** For $g = 3$, the distribution of phenotype abundances ($S_3$) is again very close to a log-normal distribution with $\mu = 5.604$ and $\sigma = 1.838$. The log-normal fit is worse than in **(a)** because there is a small bump in the right part of the distribution, where more abundant phenotypes are —due to the over representation of two-gene phenotypes (see text). **(d)** The rank distribution again shows a long tail. Only 300 phenotypes in $\mathcal{P}_3$ represent almost 99% of all genotypes. The remaining $26,000$ phenotypes are extremely rare by comparison.

## 2 Degeneracy of the genotype-phenotype map

The size of genotype space in tOYLIFE grows very quickly with the number of genes in a genotype. Since a gene in tOYLIFE is a binary string of length 20, there are $2^{20}$ different genes. A genotype is formed by choosing $g$ genes from this set with replacement (the order of genes is irrelevant). Hence, the number of genotypes with $g$ genes is $\binom{g+2^{20}-1}{g} \approx 10^{6g}/g!$ (Supplementary Material, Section 1). For $g = 2$, this number is $5.5 \times 10^{11}$, for $g = 3$, it is $1.9 \times 10^{17}$, and for larger values of $g$ it keeps growing almost exponentially. A complete exploration of these genotype spaces is well beyond our computational possibilities in general. However, using computational tricks, we have exhaustively analysed the $g = 2$ and $g = 3$ cases.

We have restricted ourselves to studying those genotypes that are able to catabolise at least one metabolite —these will be called viable genotypes. The remaining (non-viable) genotypes are unable to catabolise any metabolite. For $g = 2$, there are $1.1 \times 10^9$ viable genotypes, representing little more than 0.2% of all genotypes. For $g = 3$, this number is $1.0 \times 10^{15}$ —around 0.5% of all genotypes. In both cases, the great majority of genotypes are unable to catabolise any metabolite. But note that the space of viable genotypes is still enormous.

Among these viable genotypes, many of them catabolise exactly the same metabolites —they encode the same metabolic phenotype. For $g = 2$, there are only 775 different phenotypes, corresponding to an average of $1.4 \times 10^6$ genotypes per phenotype. For $g = 3$, there are 26,492 phenotypes, corresponding to an average of $3.8 \times 10^{10}$ genotypes per phenotype. In other words, for both $g = 2$ and $g = 3$, the degeneracy of the genotype-phenotype map is huge. From now on, we will refer to the set of phenotypes found for $g = 2$ and $g = 3$ as $\mathcal{P}_2$ and $\mathcal{P}_3$, respectively.

The distribution of phenotype abundances is highly skewed (Figure 1), similarly to what has been observed for other genotype-phenotype maps. In both cases, the distribution can be empirically fitted to a log-normal distribution. We obtained the parameters empirically from the log-transformed data (Figure 1**a** and **c**). For both $g = 2$ and $g = 3$, the rank distributions (Figure 1**b** and **d**) show a long tail, confirming that, indeed, while few phenotypes are very abundant, most of them are rare. For $g = 3$, this is especially striking, since 300 phenotypes in $\mathcal{P}_3$ represent nearly 99% of all genotypes —which means that the remaining 1% is distributed among $\sim 26,000$ phenotypes.

All phenotypes in $\mathcal{P}_2$ are also found in $\mathcal{P}_3$: we can always add a gene whose product does not fold into any protein to a viable two-gene genotype. A pertinent question, therefore, is how abundant the phenotypes belonging to $\mathcal{P}_2$ are in three-gene genotype space. We find that phenotypes in $\mathcal{P}_2$ take up 99.6% of genotypes in $g = 3$ (Supplementary Material, Section 3). This means that the 775 phenotypes in $\mathcal{P}_2$ dominate the space of phenotypes for $g = 3$. Only special combinations of three proteins and three promoters will yield most of the phenotypic diversity observed for $g = 3$. The majority of genotypes will be extensions of two-gene genotypes with a third gene that does not interfere with their function.

We can re-compute the histogram in Figure 1 taking the 775 phenotypes from $\mathcal{P}_2$ as a separate set from the remaining 25,717 phenotypes in $\mathcal{P}_3$ that are not in $\mathcal{P}_2$ (Supplementary Material, Section 3). This reveals that the small bump observed in the right part of the distribution of phenotypes in $\mathcal{P}_3$ (Figure 1**c**) is due to the phenotypes in $\mathcal{P}_2$. When we eliminate these phenotypes, the resulting distribution is much closer to a log-normal. In a sense, it is as if both sets were somehow independent: one is formed by two-gene genotypes with a third, non-interfering gene, and the other is formed by all combinations of three genes that encode new phenotypes. This influence of $\mathcal{P}_2$ phenotypes decays linearly when genotype size increases (Supplementary Material, Section 4), although they keep representing more than 80% of genotype space for $g \leq 13$.

## 3 Neutral networks in tOYLIFE

Robustness can be defined as $R = k/k_{\max}$, where $k$ is the degree of a node in a neutral network, and $k_{\max} = 20g$ is the maximum number of point-mutation neighbors. In other words, $R$ is the normalized degree of a node. We can sample genotypes for different genotype sizes, represented by $g$ (gene number), and plot the histogram of values of $R$ (Supplementary Figure 11). All the resulting distributions are unimodal, as has been observed in other genotype-phenotype maps [5, 13]. tOYLIFE genotypes become more robust as $g$ increases. In fact, there is a linear relationship between $g$ and $\langle k \rangle$, the average degree of a node in a neutral network (Figure 2**a**): $\langle k \rangle = -27.6 + 17.8g$. But $\langle R \rangle = \langle k \rangle / 20g$, so we obtain $\langle R \rangle \sim 0.891 - 1.378/g$, which is very close to the least-squares fit $\langle R \rangle = 0.895 - 1.392/g$, shown in Figure 2**b**. The linear relationship between $\langle k \rangle$ and $g$ with slope 17.8 indicates that, on average, for every gene we add to a genotype, nearly 18 out of 20 new mutations will be neutral. This implies that $\langle R \rangle$ saturates at a value close to 0.9 when $g$ increases. This result is consistent with the results of Section 2, which showed that newly added

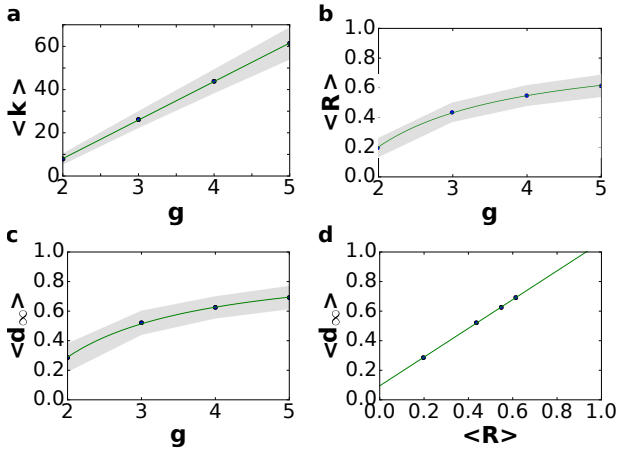


Figure 2: **Neutral networks in tOYLIFE traverse genotype space. (a)** Average degree of nodes (circles) in neutral networks (see Supplementary Figure 11 for the degree distribution) versus gene number $g$. The average degree $\langle k \rangle$ of a node grows linearly with gene number $g$, as $\langle k \rangle = -27.6 + 17.8g$ (line). **(b)** Average robustness (circles) versus gene number $g$. Robustness grows with gene number, and we can find a non-linear relationship between both variables: $\langle R \rangle = 0.895 - 1.392/g$. **(c)** There is a non-linear relationship between $g$ and $\langle d_\infty \rangle$: $\langle d_\infty \rangle = 0.965 - 1.354/g$ (line). The circles represent $\langle d_\infty \rangle$. **(d)** There is a linear relationship between $\langle d_\infty \rangle$ and the average robustness of the genotypes as obtained in **b**, given by: $\langle d_\infty \rangle = 0.094 + 0.972\langle R \rangle$ (line), very close to the $\langle d_\infty \rangle = \langle R \rangle$ fit. In all cases, the gray area encompasses two standard deviations.

genes rarely interfere with an existing phenotype. These new genes can be viewed as "junk" in the sense that they do not have any effect on metabolic function and that mutations in their sequence tend to be neutral. We will see later on that junk genes are however important, in that they enhance evolvability in tOYLIFE genotypes.

Also, taking into account that $\mathcal{P}_2$ phenotypes dominate in $\mathcal{P}_g$ for $g \leq 13$ (Supplementary Material, Section 4), we can estimate that $S_g \sim 17.8^g S_2$, so $\log S_g \sim q + g \log 17.8$, where $q$ is a constant. Combining this result with the linear relationship between $g$ and $\langle k \rangle$, we obtain for tOYLIFE the linear relationship between $\langle k \rangle$ and $\log S$, that has been observed previously for other models [13, 30, 31, 34] (but see Figure 3 for a direct verification of this relationship).

In other genotype-phenotype maps, neutral networks tend to have one giant component [18], although this is not always the case: too short RNA sequences form neutral networks that are highly disconnected [13]. Although network analysis is almost impossible for $g \geq 3$, as networks are enormous, for $g = 2$ we can perform network analyses on all 775 phenotypes exhaustively, and compute their connected components (Supplementary Material, Section 6). We observe that most phenotypes are distributed in highly fragmented neutral networks: the genotypes encoding a given phenotype form many disjoint connected components, which are typically small. Abundant phenotypes tend to have a larger number of connected components, and we can find a relatively good power-law fit between the abundance of the phenotype $S_2$ and the number of components $C$: $C = 0.25 S_2^{0.7}$ (Supplementary Material, Section 6). We also observe a huge variation in the size of connected components in $g = 2$: although more than 98% of connected components are smaller than $1,000$ nodes, some of them reach up to $\sim 10^7$ nodes. The high fragmentation of neutral networks is due to the HP model that underlies protein folding in tOYLIFE: there are no neutral mutations within proteins (Supplementary Material, Section 1, Supplementary Figure 3). Most mutations in a protein sequence yield proteins that do not fold, and few mutations yield one functional protein from another. Remember that each phenotype is defined by a list of metabolites that a genotype is able to catabolise. A given phenotype can be encoded by more than one protein combination, which will be characterised by different interactions and regulatory functions, but will catabolise the same metabolites. When a phenotype is generated by several protein combinations, it will be difficult to mutate from one combination to another, and as a consequence these combinations will usually be found in disjointed connected components. In fact, the number of connected components associated with a phenotype is positively correlated with the number of protein combinations that generate it (Supplementary Material, Section 6).

For $g \geq 2$, we can estimate the distribution of neutral networks in genotype space using neutral random walks: starting at a randomly chosen genotype, we perform a mutation on it. If the resulting mutant genotype belongs to the same neutral network —if it encodes the same phenotype— the mutation is accepted. The random walk continues when we mutate the new genotype again. If the mutant genotype does not belong to the neutral network, the mutation

is rejected, and we try to find a new neutral neighbor for the original genotype (this process will not work if the starting genotype does not have neutral neighbors, a rare case). We performed $1,000$ neutral random walks of length $10,000$ for genotype sizes $g = 2$ to $g = 5$ (Supplementary Figure 14). At each time step $t$, we computed $d_H(g_0, g_t)$, the Hamming distance (normalized number of different positions) between the original genotype $g_0$ and the genotype visited at time $t$, $g_t$. $d_H(g_0, g_t)$ is a random variable for each $t$, and so we can compute its average and standard deviation (Supplementary Figure 14). If there were no restrictions to the nodes that can be visited in a random walk, we would expect $d_H(g_0, g_t) \to 0.5$ when $t \to \infty$. In other words, if there are no restrictions, the correlation between $g_0$ and $g_t$ is lost when $t$ grows, and the distance between them tends to the value it would have, on average, if we randomly picked two genotypes from the network. Thus, the evolution of $d_H(g_0, g_t)$ is a good measure of the size and extension of neutral networks in genotype space. For $g = 2$, $\langle d_H(g_0, g_t)\rangle \sim 0.25$ when $t \to \infty$, implying that networks do not extend very far. Considering that the total genotype space has diameter 40, this means that the average distance between the initial genotype and the final one is close to 10. This is not a very high value, and it is consistent with our previous analysis showing that neutral networks in $g = 2$ tend to be fragmented and small. For $g > 2$, $\langle d_H(g_0, g_t)\rangle \sim 0.4$ when $t \to \infty$, which implies that the fragmented networks of $g = 2$ space are becoming more connected as $g$ grows, facilitating the navigability of genotype space. This suggests that neutral networks for $g > 2$ span large fractions of genotype space, a result consistent with other genotype-phenotype maps.

A different way to estimate the diameter of a neutral network is to perform neutral random walks in which we force $d_H(g_t, g_{t+1}) > d_H(g_{t-1}, g_t)$. That is, in addition to imposing that a mutation is neutral in order to accept it, we also require it to increase the distance to the original genotype. More specifically, the process is computed as follows. We randomly choose a genotype, and perform mutations on it, increasing the distance every time step, until this distance can increase no longer —if, after a large number of trials, we cannot find a neutral mutant that is farther apart from the original genotype, we stop the process. We will denote the final distance obtained in such random walks by $d_\infty$. For $g = 2$ and $g = 3$ we randomly sampled $10,000$ genotypes, whereas for $g = 4$ and $g = 5$ we sampled $1,000$ genotypes (Figure 2**c**, Supplementary Figure 15). Consistent with previous results, random walks did not get very far for $g = 2$, reaching an average final distance $\langle d_\infty\rangle \sim 0.28$. For $g > 2$, the final distance $d_\infty$ increases. This result confirms the previous observation that navigability in these genotype spaces is enhanced. For $g = 3$, $\langle d_\infty\rangle$ is a little over 0.5, while for $g = 4$ and $g = 5$ it reaches 0.6 and 0.7, respectively. In fact, the growth of $\langle d_\infty\rangle$ with $g$ is very similar to the growth of $\langle R\rangle$ obtained in Figure 2**b**. In that case, we had $\langle R\rangle = 0.895 - 1.392/g$. Here it is $\langle d_\infty\rangle = 0.965 - 1.354/g$ (Figure 2**c**). Unsurprisingly, the similarity of the fits implies a linear relationship between $\langle d_\infty\rangle$ and $\langle R\rangle$: $\langle d_\infty\rangle = 0.094 + 0.972\langle R\rangle$ (Figure 2**d**), very close to the identity function. This result has several implications. First, as $g$ grows, neutral networks are more and more connected, and they span larger fractions of genotype space. It is easier to get from one extreme of the genotype network to the other without changing the phenotype. Secondly, this increased connectivity is due to the increase in robustness: the robustness of a genotype is a good predictor for the size of the connected component it belongs to. This can be easily explained in light of our previous discussion on robustness. Adding a new gene to a genotype will endow the latter with an average of 18 new neutral mutations with which to explore genotype space (Figure 2**a**). Because the new gene will not interfere with the phenotype with a high probability, it follows that we can mutate most of its nucleotides, one by one, getting farther away from the original genotype. In other words, new

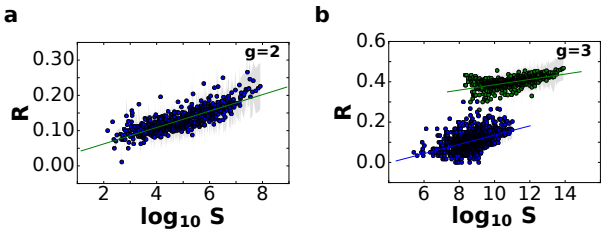


Figure 3: **Phenotypic robustness is linearly related to the logarithm of phenotype abundance.** For $g = 2$ and $g = 3$, we sampled $10^7$ genotypes and computed their robustness. Then we assigned each of them to their corresponding phenotypes, and estimated phenotypic robustness, the average robustness for all genotypes encoding a given phenotype (see text). **(a)** Phenotypic robustness in $g = 2$ versus the logarithm of phenotype abundance. The green line represents the power-law relationship $R_p = 1.037\, S_2^{0.023}$. **(b)** For $g = 3$, we separated those phenotypes belonging to $\mathcal{P}_2$ (green circles) from the rest (blue circles). Both sets show a power law relationship between phenotypic robustness and phenotypic abundance: $R_p = 1.790\, S_3^{0.013}$ for phenotypes in $\mathcal{P}_2$ (green line), and $R_p = 0.805\, S_3^{0.023}$ for the remaining $25,717$ phenotypes (blue line). Gray area encompasses two standard deviations.

genes in toyLIFE allow for increased navigability of genotype space, because they are mostly junk genes. As we will see later on, this property will have important consequences for evolvability.

The fact that robustness is a good predictor for the size of a genotype's connected component can be combined with the positive correlation between the logarithmic abundance of a phenotype and the size of its largest connected component (Supplementary Material, Section 6) to deduce the linear relationship between the logarithm of phenotype abundance and phenotypic robustness, giving yet another heuristic argument for this relationship. We now turn to compute it explicitly.

Phenotypic robustness is defined as the average of genotypic robustness for all genotypes encoding a phenotype $\mathcal{P}_i$, that is

$$R_{\mathcal{P}_i} = \frac{1}{|\mathcal{P}_i|} \sum_{g \in \mathcal{P}_i} R_g,$$

where $|\mathcal{P}_i|$ is the number of genotypes encoding $\mathcal{P}_i$. For $g = 2$ and $g = 3$ we sampled $10^7$ genotypes and computed their robustness. We then assigned each genotype to its corresponding phenotype and averaged the values of robustness for all genotypes encoding each phenotype. Note that this procedure samples abundant phenotypes more often. For $g = 2$, we find a good fit to a linear relationship between the logarithm of phenotype abundance and estimated phenotypic robustness (Figure 3**a**). For $g = 3$, we identified those phenotypes belonging to $\mathcal{P}_2$ (being the most abundant, they were sampled the most) in green, and the rest in blue (Figure 3**b**). The Figure shows separate relationships between the logarithm of phenotype abundance and phenotypic robustness. The two sets of phenotypes cluster in two different groups, confirming once more the idea that these two sets are qualitatively different. Phenotypes belonging to $\mathcal{P}_2$ are much more robust, as a result of them having one spare junk gene.

# 4 Robustness and position in genotype

Instead of considering the degree of a node in a neutral network, we can focus on the neutrality of a given position of the genotype. A genotype formed by $g$ genes can be thought of as a binary string of length $20g$ —remember that genes in toyLIFE have 20 nucleotides, the first four forming the promoter region and the remaining 16 constituting the coding region. For a given sequence, the position $i = 1, \ldots, 20g$ can either be neutral or not —that is, when we mutate that position, we can get a new genotype with the same phenotype or not. We can thus define the random variable

$$r_i = \begin{cases} 1 & \text{if } i \text{ is a neutral position,} \\ 0 & \text{otherwise.} \end{cases}$$

Because $r_i$ is a random variable, we can sample it and estimate its mean. Differences between positions may yield insights into the details of the genotype-phenotype correspondence in toyLIFE —i.e. some positions may always be

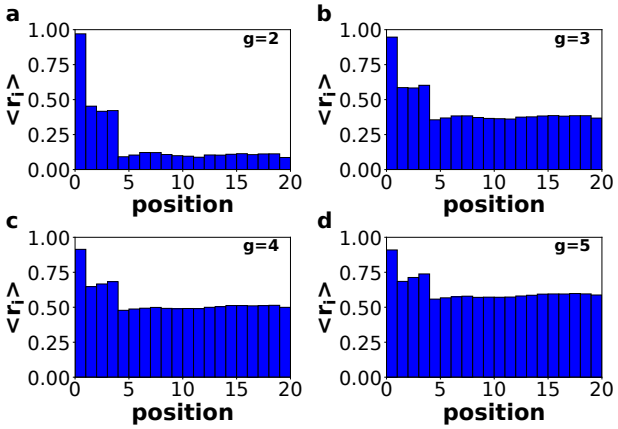


Figure 4: **Different positions in the genome have different neutralities.** We sampled $10^7$ genotypes for $g = 2$ and $g = 3$ and $10^3$ genotypes for $g = 4$ and $g = 5$, and measured $r_i$ for $i = 1, \ldots, 20$. For each $i$ we then computed $\langle r_i \rangle$ and plotted them versus genomic position. Note the high robustness of the first position in the promoter region, and the low robustness in the coding regions.

neutral, or always constrained. This is what we have done in Figure 4, for genotype sizes $g = 2$ to $g = 5$. We sampled $10^7$ genotypes for $g = 2$ and $g = 3$, and $10^3$ genotypes for $g = 4$ and $g = 5$, and computed $r_i$ for every $i = 1, \ldots, 20g$ and every genotype. The order of genes does not matter in tOYLIFE by construction —implying $\langle r_i \rangle = \langle r_{i+20h} \rangle$, for any $h = 0, 1, \ldots, g-1$—, so we are interested in the values of robustness for each gene. This is why in Figure 4 we only show the average values $\langle r_i \rangle$ for $0 \leq i < 20$. Note that promoter regions tend to be more robust than coding regions. This is partly due to the lack of robustness in the version of the HP model that underlies protein folding in tOYLIFE (Supplementary Material, Section 1, Supplementary Figure 3). However, note that the superposition of regulatory and metabolic levels of the phenotype makes the average robustness of coding regions grow, in spite of the non-robust protein folding model. For $g = 4$ and $g = 5$, the average robustness in these regions reaches values as high as 0.5 (remember that these genotypes tend to have junk genes that increase overall robustness as $g$ grows).

Inside the promoter region, which only affects gene regulation, the first position is particularly robust. This means that the regulatory changes it induces have no phenotypic effect at the metabolic level. This may be due to two reasons: either changes in the first position of the promoter region do not affect the regulatory function —the temporal pattern of gene expression determined by the interactions among proteins— or changes in the regulatory function rarely alter the metabolic phenotype. We performed the following simple test of these hypotheses. For each position in the promoter region, we sampled $10^4$ genotypes of size $g = 3$. We then mutated that position and computed the new regulatory function and the new metabolic phenotype. From all $10^4$ mutations in the first position, 40% were neutral in both the regulatory and the metabolic sense, 54% affected the regulatory function but did not affect the metabolic phenotype, and the remaining 6% changed both —this means that the robustness for the first position in this sample was 94%. For the rest of the positions, 27% of the mutations did not alter either the regulatory function or the metabolic phenotype, 32% changed regulation but not metabolism, and 41% changed both. This corresponds to a robustness of 59%, consistent with what we observed in Figure 4. In other words, for the first position only 9% of the mutations that affected regulation had any effect on the metabolic phenotype. In addition, 40% of mutations did not affect the regulatory function at all. For the rest of the positions, however, the number of mutations that altered regulation, 73%, was higher. Among these, roughly 55% had an effect on phenotype as well. So both reasons posited above are at work: not only is the number of mutations affecting regulatory function lower in the first position of the promoter region, but when these mutations do alter the regulatory function, they rarely change the phenotype.

The lower robustness of coding regions, compared to promoter regions, is correlated with a higher evolvability, as will be discussed in the next Section.

# 5 Accessibility and Evolvability

So far, we have limited our discussion of the properties of the genotype-phenotype map in tOYLIFE to the abundance of phenotypes and the organisation of their neutral networks, without paying any attention to the connections between different phenotypes. In this section we will focus on the latter question, which amounts to studying evolvability, or how accessible phenotypes are.

The neutral networks of different phenotypes tend to be highly interwoven in most computational genotype-phenotype maps, so that connections between them are very common. The Vienna RNA group described a property of RNA neutral networks called shape space covering [7, 9]. It implies that one can find most common phenotypes a few mutations away from any given genotype. We checked for the existence of this property in tOYLIFE. We sampled 100 genotypes for $g = 2$ and $g = 3$ and computed the phenotypes of all neighbors at distances 1 to 8. We observed how many of the 300 most common phenotypes appeared in this set of neighbors. The results are shown in Figure 5**a** and **b**. For both $g = 2$ and $g = 3$ and most sampled genotypes, the number of phenotypes discovered after 8 mutations was close to 300. This implies that tOYLIFE also shares the shape space covering property: most phenotypes are just a few mutations away from any given phenotype. Observe, however, that for $g = 2$ this means a higher relative distance —remember that the diameter of this network is 40— and that the number of phenotypes discovered at that distance is lower by comparison.

Shape space covering means that phenotypes are easily accessible from each other through a few number of mutations. A relevant detail in the metabolic genotype-phenotype map in tOYLIFE is that this accessibility is due only to mutations in proteins. If we mutate only the promoters, the number of visited phenotypes is never larger than 2, regardless of the distance. For $g = 2$, we can give a clear explanation to this peculiarity: of the 135,318 pairs of proteins that yield a metabolic function, only 16 yield two different metabolic phenotypes when combined with different promoters. The rest are able to generate only one phenotype. Changing only the promoters will not affect

the metabolic function, and will not help in finding new phenotypes. This is consistent with the high robustness of the promoter region.

Another way to study evolvability is to compute the connections between different phenotypes directly. We say that two phenotypes are connected if at least two genotypes encoding each phenotype are one point mutation away from each other. We can then create a network of phenotypes, whose nodes will be the phenotypes themselves, and the edges the connections between them. This network of phenotypes is undirected and weighted —the weight of an edge between two phenotypes is the sum of all point mutations connecting two genotypes encoding each phenotype. This network admits self-loops, whose weight is twice the number of edges connecting genotypes encoding a single phenotype —in other words, it is the sum of the degrees of all the nodes encoding that phenotype, which is proportional to the phenotype's robustness. For $g = 2$, where we can compute the whole network of genotypes with their corresponding phenotypes, we can build this phenotype network exhaustively. The network is formed by a giant component that includes 767 phenotypes. We also find six additional tiny components, five of them with just one phenotype and the remaining one with three phenotypes (Supplementary Material, Section 8). Thus, for $g = 2$, some phenotypes will be unreachable by point mutations from other phenotypes. For $g = 3$, we cannot build the phenotype network exhaustively, but resorting to a numerical approximation using random walks (Supplementary Material, Section 8), we can estimate the network of connections between the 775 phenotypes in $\mathcal{P}_2$ — in order to study how the addition of one gene alters the connections between these phenotypes. The results show that all phenotypes in $\mathcal{P}_2$ now belong to one giant component (Supplementary Material, Section 8). The number of connections between phenotypes has greatly increased as well. This is again due to the additional junk genes. They do not only increase robustness, but also allow for increased connections between phenotypes.

This increased connectivity can also be measured in an alternative way. In previous work [5], evolvability has been estimated as the number of new phenotypes discovered in a neutral random walk along a neutral network. In Figure 5**c** and **d** we have performed such an analysis for $10,000$ genotypes for $g = 2$ and $g = 3$. The results show that evolvability is much higher for $g = 3$. While the number of discovered phenotypes almost stops growing for $g = 2$, it grows quickly in $g = 3$, and to a much higher value than for $g = 2$. Again, this is due to the higher average number of connections between phenotypes for $g = 3$ (Supplementary Material, Section 8).

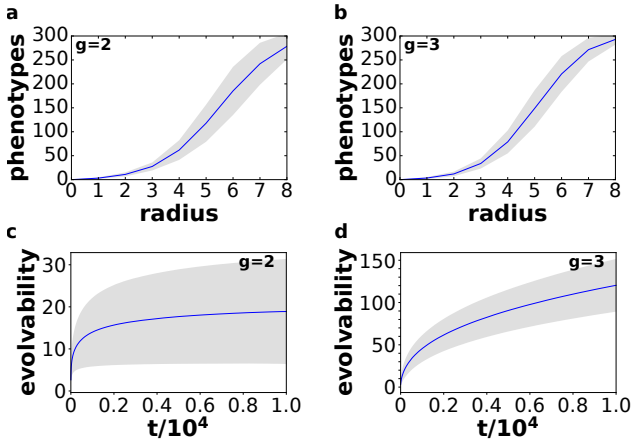


Figure 5: **Evolvability in toyLIFE.** A genotype-phenotype map has the shape space covering property if, given a phenotype, we only need to explore a small radius around a sequence encoding that phenotype in order to find the most common phenotypes. We tested this property in toyLIFE by sampling 100 genotypes for $g = 2$ (**a**) and $g = 3$ (**b**), and counting how many of the 300 most common phenotypes appeared in a radius of distance 8 around a given genotype. The results are consistent with shape space covering. For $g = 2$ (**c**) and $g = 3$ (**d**), we also measured the cumulative number of phenotypes in the neighborhood of a neutral random walk —evolvability— for $10,000$ genotypes. Evolvability is much higher for $g = 3$. Lines show average values, while gray areas encompass two standard deviations.

# 6 Discussion and Conclusions

Throughout this article, we have explored the properties of the metabolic genotype-phenotype map in tOYLIFE. This map is highly degenerate, with many more genotypes than phenotypes, and large neutral networks traversing genotype space. The distribution of phenotype abundances is very heterogeneous, and larger phenotypes tend to be more robust. Common phenotypes are easily accessed from each other, and large neutral networks allow for a fast exploration of phenotype space.

All of these properties have been described in other genotype-phenotype maps [7–31, 33, 34]. This is somewhat striking, given that the genotype-phenotype map in tOYLIFE is more complex than the rest of these models. It is the only model that incorporates intermediate levels of phenotypic expression: genes are first translated into proteins, which fold and interact with each other, generating complex regulatory networks that will determine the metabolic capacities of a genotype. And yet the main properties shared by the rest of genotype-phenotype maps appear here as well.

Two particular results stand out. The first is the log-normal distribution of phenotype abundances, which has also been observed in RNA [14] and predicted for simple combinatorial genotype-phenotype maps [34]. The second is the positive correlation between phenotypic robustness and the logarithm of phenotype abundance, which has also been described before [13, 30, 31, 34]. The fact that these two relationships (as well as other phenomena, such as shape space covering) are so widespread points to a general property of these maps, which must be related to combinatorics and network theory. Previous work [34] has shown that, when the abundance of a phenotype can be inferred from the genotype sequence in simple genotype-phenotype maps, we can use combinatorial arguments to explain the appearance of a log-normal distribution and the linear relationship between phenotypic robustness and the logarithm of phenotype abundance. These arguments would explain the presence of these properties in the case of RNA, but do not seem to be easily translatable to tOYLIFE. We need to devote more efforts into understanding this seemingly general property of genotype-phenotype maps.

The high robustness and evolvability of the metabolic phenotype in tOYLIFE, particularly when genome size increases, is remarkable because our model is built on a particularly non-robust, non-evolvable version of the HP protein-folding model. Proteins in tOYLIFE are sequences of 16 amino-acids that fold on a $4 \times 4$ lattice. One protein is defined by its perimeter and its folding energy: this is a very different definition of protein than the one used in most versions of the HP model [15–20], which define a protein by its folded structure. As we have already mentioned, there are no neutral networks in the protein space in tOYLIFE, and evolvability is very limited. However, when these proteins are paired to interact with each other, generating regulatory networks and performing metabolic functions, the resulting genotypes are robust and evolvable. So it would seem that adding levels of expression to a phenotype enhances both robustness and evolvability. This hypothesis could be tested, for instance, by adding another level of complexity to the phenotype in tOYLIFE and checking how both of these properties are changed. If there is indeed a relationship between complexity and both robustness and evolvability, this result would suggest that more complex genotype-phenotype maps could have an evolutionary advantage.

Alternatively, we could change the folding process in tOYLIFE and allow for promiscuous proteins —proteins that can fold in different shapes with the same energy [41]. This would increase connectivity in protein space and would affect the levels of robustness and evolvability at the metabolic level. Determining the extent of this change would give us some insight into the relationship between different levels of phenotypic expression.

On a related note, our results show that adding genes to tOYLIFE genotypes increases both robustness and evolvability. Neutral networks of two-gene genotypes are not very navigable, reaching only a small portion of genotype space, and connecting with a small number of adjacent phenotypes. However, adding a new gene to the genotype changes everything: now phenotypes are easily accessed from each other and neutral networks span genotype space. Robustness, as we haven seen, keeps growing with genome size. The explanation behind this fact is that most of the new genes will not alter the metabolic phenotype, and will act as junk genes. However, they can mutate without restriction, enhancing the navigability of a neutral network. Increased navigability allows for increased connectivity between phenotypes, thus enhancing evolvability [42]. In other words, junk genes have creative potential, in the sense that they allow populations to explore a given neutral network, and then encounter new, unexplored phenotypes. This is interesting because it extends the usefulness of redundancy in complex genomes [43, 44] to include seemingly inert elements, whose only function is to increase robustness and evolvability. It is also reminiscent of the abundance of introns and non-coding DNA in eukaryotic genomes [45]: if this non-functional DNA also enhances robustness and evolvability in living cells, this would suggest new arguments for the maintenance of junk DNA.

Finally, the appearance of junk genes is possibly a result of the fact that interactions in tOYLIFE are limited to

be pairwise. There are neither trimers nor tetramers in tQYLIFE, only dimeric proteins. Only one protein or dimer can interact with a metabolite at a given moment, and so on. As a consequence, when a new gene is added to a two-gene genotype that performs a metabolic function, it will have little potential to create new functions. In fact, one would expect the opposite: that adding a new gene would disrupt the existing interactions, thus yielding a non-viable metabolic phenotype. However, this is not what we observe. When adding new genes to two-gene genotypes, most genotypes keep their original function. We will need to perform a more detailed exploration of this phenomenon in order to clarify the reasons behind it.

# References


[1] Alberch P. From genes to phenotype: dynamical systems and evolvability. Genetica. 1991;84:5–11.

[2] Eldredge N, Gould SJ. Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf TJM, editor. Models in Paleobiology. San Francisco: Freeman Cooper; 1972. p. 82–115.

[3] Gould SJ, Eldredge N. Punctuated equilibria: the tempo and mode of evolution reconsidered. Paleobiology. 1977;3:115–151.

[4] Maynard Smith J, Burian R, Kauffman S, Alberch P, Campbell J, Goodwin B, et al. Developmental constraints and evolution: a perspective from the Mountain Lake conference on development and evolution. Q Rev Biol. 1985;p. 265–287.

[5] Wagner A. The Origins of Evolutionary Innovations. Oxford University Press; 2011.

[6] Nei M. Mutation-driven Evolution. OUP Oxford; 2013.

[7] Schuster P, Fontana W, Stadler PF, Hofacker IL. From sequences to shapes and back: A case study in RNA secondary structures. Proc Roy Soc London B. 1994;255:279–284.

[8] Grüner W, Giegerich R, Strothmann D, Reidys C, Weber J, Hofacker IL, et al. Analysis of RNA sequence structure maps by exhaustive enumeration. I. Neutral networks. Monatshefte f Chemie. 1996;127:355–374.

[9] Grüner W, Giegerich R, Strothmann D, Reidys C, Weber J, Hofacker IL, et al. Analysis of RNA sequence structure maps by exhaustive enumeration II. Structures of neutral networks and shape space covering. Monatshefte f Chemie. 1996;127:375–389.

[10] Huynen MA. Exploring phenotype space through neutral evolution. J Mol Evol. 1996;43:165–169.

[11] Fontana W, Schuster P. Continuity in evolution: on the nature of transitions. Science. 1998;280:1451–1455.

[12] Jörg T, Martin OC, Wagner A. Neutral network sizes of biological RNA molecules can be computed and are not atypically small. BMC Bioinformatics. 2008;9:1.

[13] Aguirre J, Buldú JM, Stich M, Manrubia SC. Topological structure of the space of phenotypes: the case of RNA neutral networks. PLoS ONE. 2011;6:e26324.

[14] Dingle K, Schaper S, Louis AA. The structure of the genotype–phenotype map strongly constrains the evolution of non-coding RNA. Interface focus. 2015;5:20150053.

[15] Lau KF, Dill KA. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. Macromolecules. 1989;22:3986–3997.

[16] Lipman DJ, Wilbur WJ. Modelling neutral and selective evolution of protein folding. Proc Roy Soc London B. 1991;245:7–11.

[17] Li H, Helling R, Tang C, Wingreen N. Emergence of preferred structures in a simple model of protein folding. Science. 1996;273:666–669.

[18] Bornberg-Bauer E. How are model protein structures distributed in sequence space? Biophys J. 1997;73:2393.

[19] Bastolla U, Roman HE, Vendruscolo M. Neutral evolution of model proteins: diffusion in sequence space and overdispersion. J Theor Biol. 1999;200:49–64.

[20] Irbäck A, Troein C. Enumerating designing sequences in the HP model. J Biol Phys. 2002;28:1–15.

[21] Ciliberti S, Martin OC, Wagner A. Innovation and robustness in complex regulatory gene networks. Proc Natl Acad Sci USA. 2007;104:13591–13596.

[22] Cotterell J, Sharpe J. An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients. Mol Sys Biol. 2010;6:425.

[23] Payne JL, Moore JH, Wagner A. Robustness, evolvability, and the logic of genetic regulation. Artif Life. 2014;20:111–126.

[24] Jiménez A, Cotterell J, Munteanu A, Sharpe J. Dynamics of gene circuits shapes evolvability. Proc Natl Acad Sci USA. 2015;112:2103–2108.

[25] Rodrigues JFM, Wagner A. Evolutionary plasticity and innovations in complex metabolic reaction networks. PLoS Comp Biol. 2009;5:e1000613.

[26] Rodrigues JFM, Wagner A. Genotype networks, innovation, and robustness in sulfur metabolism. BMC Sys Biol. 2011;5:39.

[27] Wagner A, Andriasyan V, Barve A. The organization of metabolic genotype space facilitates adaptive evolution in nitrogen metabolism. J Mol Biochem. 2014;3.

[28] Hosseini SR, Barve A, Wagner A. Exhaustive Analysis of a Genotype Space Comprising 10 15 Central Carbon Metabolisms Reveals an Organization Conducive to Metabolic Innovation. PLoS Comput Biol. 2015;11:e1004329.

[29] Johnston IG, Ahnert SE, Doye JP, Louis AA. Evolutionary dynamics in a simple model of self-assembly. Phys Rev E. 2011;83:066105.

[30] Greenbury SF, Johnston IG, Louis AA, Ahnert SE. A tractable genotype-phenotype map modelling the self-assembly of protein quaternary structure. J R Soc Interface. 2014;6:20140249.

[31] Greenbury SF, Schaper S, Ahnert SE, Louis AA. Genetic correlations greatly increase mutational robustness and can both reduce and enhance evolvability. PLoS Comput Biol. 2016;12:e1004773.

[32] Arias CF, Catalán P, Manrubia S, Cuesta JA. toyLIFE: a computational framework to study the multi-level organisation of the genotype-phenotype map. Sci Rep. 2014;4:7549.

[33] Greenbury S, Ahnert S. The organization of biological sequences into constrained and unconstrained parts determines fundamental properties of genotype–phenotype maps. J R Soc Interface. 2015;12:20150724.

[34] Manrubia S, Cuesta JA. Distribution of genotype network sizes in sequence-to-structure genotype–phenotype maps. J R Soc Interface. 2017;14:20160976.

[35] Fortuna MA, Zaman L, Ofria C, Wagner A. The genotype-phenotype map of an evolving digital organism. PLoS computational biology. 2017;13(2):e1005414.

[36] Schaper S, Louis AA. The arrival of the frequent: how bias in genotype-phenotype maps can steer populations to local optima. PLoS ONE. 2014;9:e86635.

[37] Maynard Smith J. Natural selection and the concept of a protein space. Nature. 1970;225:563–564.

[38] Ciliberti S, Martin OC, Wagner A. Robustness can evolve gradually in complex regulatory gene networks with varying topology. PLoS Comput Biol. 2007;3:e15.

[39] Dall'Olio GM, Bertranpetit J, Wagner A, Laayouni H. Human genome variation and the concept of genotype networks. PLoS ONE. 2014;9:e99424.

[40] Barve A, Wagner A. A latent capacity for evolutionary innovation through exaptation in metabolic systems. Nature. 2013;500:203–206.

[41] Khersonsky O, Tawfik DS. Enzyme promiscuity: a mechanistic and evolutionary perspective. Ann Rev Biochem. 2010;79:471–505.

[42] Wagner A. Robustness and evolvability: a paradox resolved. Proceedings of the Royal Society of London B: Biological Sciences. 2008;275:91–100.

[43] Daniels BC, Chen YJ, Sethna JP, Gutenkunst RN, Myers CR. Sloppiness, robustness, and evolvability in systems biology. Curr Opin Biotech. 2008;19:389–395.

[44] Whitacre J, Bender A. Degeneracy: a design principle for achieving robustness and evolvability. Journal of Theoretical Biology. 2010;263(1):143–153.

[45] Doolittle WF. Is junk DNA bunk? A critique of ENCODE. Proceedings of the National Academy of Sciences. 2013;110(14):5294–5300.